

The TissueNet database of human tissue protein–protein interactions

Ruth Barshir¹, Omer Basha¹, Amir Eluk¹, Ilan Y. Smoly², Alexander Lan² and Esti Yeger-Lotem^{1,3,*}

¹Department of Clinical Biochemistry, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel,

²Department of Computer Science, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel and

³National Center for Biotechnology in the Negev, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel

Received August 28, 2012; Revised September 22, 2012; Accepted October 30, 2012

ABSTRACT

Knowledge of protein–protein interactions (PPIs) is important for identifying the functions of proteins and the processes they are involved in. Although data of human PPIs are easily accessible through several public databases, these databases do not specify the human tissues in which these PPIs take place. The TissueNet database of human tissue PPIs (<http://netbio.bgu.ac.il/tissuenet/>) associates each interaction with human tissues that express both pair mates. This was achieved by integrating current data of experimentally detected PPIs with extensive data of gene and protein expression across 16 main human tissues. Users can query TissueNet using a protein and retrieve its PPI partners per tissue, or using a PPI and retrieve the tissues expressing both pair mates. The graphical representation of the output highlights tissue-specific and tissue-wide PPIs. Thus, TissueNet provides a unique platform for assessing the roles of human proteins and their interactions across tissues.

INTRODUCTION

Proteins act through interactions with other molecules, and knowledge of these interactions can help identify the functions of proteins and their involvement in various processes during health and disease (1,2). Owing to their importance, many experimental methods were developed and applied to detect physical protein–protein interactions (PPIs). These methods include *in vivo* assays such as affinity-based methods, as well as high-throughput *in vitro* assays such as protein arrays (3). The various types of PPIs have been recorded in several PPI databases

that serve a large community of researchers interested in revealing the roles of proteins in various organisms [e.g. (4–7)].

Here, we focus on the human PPIs. Unlike unicellular organisms such as yeast, the human body is composed of many tissues, each expressing a distinct set of proteins [e.g. (8,9)]. Consequently, many proteins have varying PPI partners across tissues (10,11). A tissue-sensitive view of PPIs is therefore important for assessing the various roles of human proteins. However, this view is not readily available in current databases: PPI databases do not hold tissue information, potentially because many PPIs were detected using *in vitro* assays, regardless of specific tissues. Likewise, protein-centric databases report the expression of a protein across tissues, but do not associate its PPIs with a specific tissue [e.g. (12,13)].

TissueNet was designed to provide a tissue-sensitive view of human PPIs. To this end, we integrated current data of human PPIs from four major PPI databases with data of tissue expression of genes and proteins. The tissue expression data were assembled from three extensive resources, each based on a different profiling technique, and resulted in mapping of proteins to 16 main human tissues. We then associated each PPI with the subset of tissues that express both pair mates, thus leaving out tissues that are less likely to contain the PPI.

Users can query TissueNet using a protein and retrieve its PPI partners per tissue, or using a PPI and retrieve the tissues expressing both pair mates. The graphical network output distinctly highlights tissue-specific and tissue-wide proteins and PPIs. TissueNet also provides auxiliary information regarding proteins and interactions, such as protein expression levels across tissues, PPI detection methods and gene ontology (GO) annotations. We believe that TissueNet will be of great help to researchers interested in the various roles of human protein and interactions across tissues.

*To whom correspondence should be addressed. Tel/Fax: +972 8 6428675; Email: estiy@bgu.ac.il

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

© The Author(s) 2012. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits non-commercial reuse, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com.

METHODS

Expression data sources

Data of mRNA levels across tissues, denoted GNF (8), were downloaded from BioGPS (14), and all genes with intensity value >100 in a tissue were considered as expressed (15). Data of protein expression across tissues, denoted Human Protein Atlas (HPA) (16), were limited to proteins with a positive expression value (ranked: 1, 2, 3, 4) in a tissue. Proteins were further filtered by imposing stringent thresholds on the reliability and validity of their antibodies. Specifically, when antibody-reliability score was available, a medium or high was required; otherwise at least one supportive and no negative validity scores were required. RNA-seq data from Illumina Body Map 2.0 (17) were filtered for genes with at least 1 RPKM. All analyses were limited to proteins and protein-coding genes only, and these were mapped to their Ensembl gene ids using BioMart (18). Supplementary Table S1 presents an overview of the objects and tissues measured in each resource.

Consolidation of expression and tissue data

As RNA-seq data covered the largest number of genes per tissue, we based our analysis on the 16 main human tissues profiled with RNA-seq. As GNF and HPA profiled sub-parts of these tissues, we manually consolidated the various tissue sub-parts according to the consolidation scheme given in Supplementary Table S2.

RESULTS

TissueNet synergizes between large-scale data of PPI and expression profiling across tissues to create a unique database of tissue-associated PPIs. Below, we describe the construction of TissueNet and its usage.

TissueNet construction

In TissueNet, a PPI is associated with the tissues expressing both pair mates. This was achieved by integrating recent data of PPIs with extensive data of gene and protein expression across 16 main human tissues. Specifically, experimentally detected human PPIs were assembled for four major PPI databases, including BIOGRID (4), DIP (5), IntAct (6) and MINT (7). These data amounted to 67 439 PPIs between 11 225 human proteins. Data of gene and protein expression across tissues were assembled from three major resources: the GNF data set of Su *et al.* (8) based on profiling using DNA microarrays, the HPA based on protein immunohistochemistry measurements (16) and the Illumina Body Map 2.0 based on RNA-seq measurements (17). These data were subject to stringent thresholds (see Methods). As RNA-seq identified the largest number of genes per tissue (9,11), we focused on the 16 main human tissues profiled by RNA-seq. The 16 tissues were as follows: adipose, adrenal, brain, breast, colon, heart, kidney, liver, lung, lymph node, ovary, prostate, skeletal muscle, testis, thyroid and white blood cells. We then combined data of these tissues from the three resources as follows: We associated a PPI with a

tissue if each pair mate was found to be expressed in that tissue according to at least one resource. For tissues consisting of multiple sub-parts, such as brain, we associated a PPI with the tissue only if the two pair mates were detected in similar or closely related sub-parts of that tissue (see Methods). Consequently, 59 640 PPIs were associated with at least 1 of the 16 tissues.

TissueNet usage

Users query TissueNet using a protein and retrieve its PPI partners per tissue, or using a PPI and retrieve the tissues expressing both pair mates. The output of TissueNet includes a network view using a Cytoscape Web plug-in (19), as well as a textual output and a context menu for consecutive user queries. Below, we describe the input, output and context menu in more detail.

Querying TissueNet

TissueNet supports protein and PPI queries that use Ensembl gene ID, Entrez gene ID or wiki names. The full lists of proteins that are present in TissueNet are accessible from TissueNet homepage. In addition to the protein and PPI queries, TissueNet also offers a ‘sample protein’ query that retrieves the tissue PPIs of the human protein TP53 and a ‘random protein’ query that retrieves the tissue PPIs of a randomly selected protein from TissueNet.

TissueNet output

TissueNet output includes a graphical network view and textual information of the output proteins and PPIs (Figure 1). The graphical network view depicts proteins as network nodes and PPIs as network edges while highlighting their tissue specificity: tissue-specific proteins, which are expressed in at most three tissues, are coloured orange, and tissue-wide proteins, which are expressed in 14–16 tissues, are coloured blue. Other proteins are coloured grey. The width of a PPI edge increases with the number of tissues associated with it.

The textual information is divided into tabs, each relating to a different type of data regarding the output. The ‘Properties’ tab lists the tissues associated with each protein or PPI and also specifies the detection methods and source databases for each PPI. The ‘Gene ontology’ (GO) tab provides the GO annotations of proteins selected in the network view (20). The ‘Chemicals’ tab lists chemicals known to interact with selected proteins, which might suggest candidates for experimental manipulation (21). Lastly, the ‘Tissues’ tab lists per tissue the subset of proteins that are expressed in that tissue, along with the reporting expression resource and measured expression value. Thus, the various tabs enable users to assess the reliability of PPIs and of tissue associations, and to learn more about the selected proteins and PPIs.

The context menu

The context menu is activated by a right mouse click over the output network view. This menu provides an interface for several user options, including changing the graph layout, exporting the output, removing proteins from the network and adding the PPIs of selected proteins

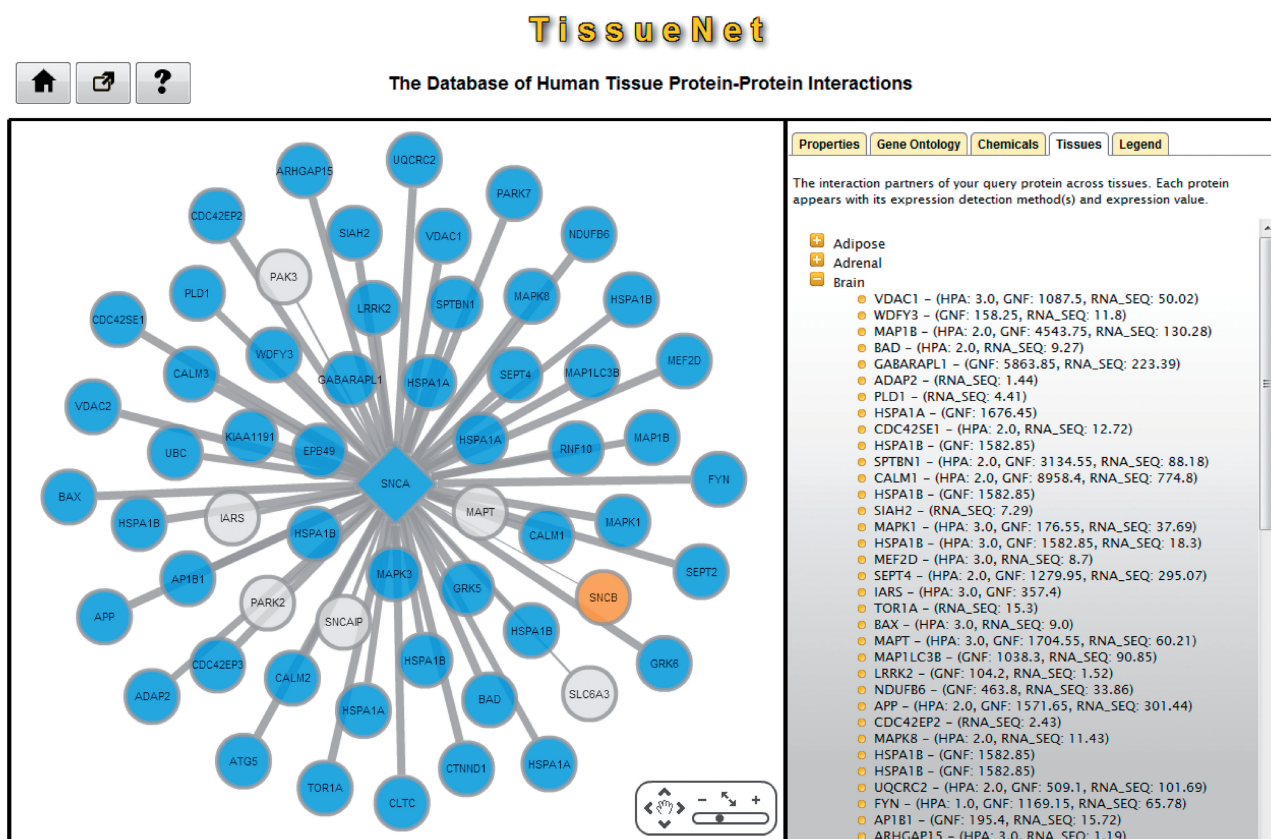


Figure 1. TissueNet presentation of the PPIs of the human protein alpha-synuclein (SNCA) in brain. The network view on the left shows SNCA as the diamond-shaped node in the middle, circled by its PPI partners in brain. The network view distinguishes between partners expressed in ≥ 14 tissues (coloured blue), partners expressed in at most three tissues (SNCB, coloured orange) and other partners (coloured grey). The Tissues tab on the right shows the measured expression levels in brain of SNCA's PPI partners.

to the current network view, thereby gradually expanding the network.

DISCUSSION

The TissueNet database of human tissue PPIs at <http://netbio.bgu.ac.il/tissuenet/> offers a unique service that associates experimentally detected PPIs with tissues that express both pair mates. To construct TissueNet, we gathered tissue-profiling data from three large resources, each using a different technique to measure expression within a tissue. By applying a stringent threshold for calling expression within a resource and uniting these data, we created a broad, high-confidence data set of gene and protein expression across 16 main human tissues. The integration of these data with data of PPIs resulted in tissue associations for 88% of the PPIs.

Although co-expression in a tissue is a necessary condition for a PPI to occur, it is important to clarify that it does not validate the PPI within a tissue. PPIs are probabilistic events that depend on additional parameters, such as the cellular localization, post-translational modifications and concentrations of the interacting proteins, which are yet unavailable for most PPIs. For these reasons, the common practice in the field of PPI

mapping is to consider experimentally identified PPIs as potential PPIs [e.g. (2)]. TissueNet further limits these potential PPIs by tissue co-expression, thereby filtering out PPIs from tissues that do not express both pair mates. To judge the filtering power of TissueNet, we calculated the fraction of mapped PPI tissue associations as follows: we summed the number of different tissue associations per PPI across all PPIs (729 675 PPI tissue associations). We then divided it by the total number of potential tissue associations for all PPI, i.e. the total number of PPIs multiplied by 16 tissues (1 079 024 PPI tissue associations). This resulted in 67.6%, meaning that TissueNet filtered out a considerable 32.4% of the potential PPI tissue associations. To help users further assess the reliability of specific PPIs, TissueNet provides for each PPI its detecting methods and the tissue expression levels of its proteins. To summarize, TissueNet is a novel platform for obtaining a tissue-sensitive view of human PPIs that can greatly facilitate the functional assessment of human proteins and their PPIs across tissues.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1 and 2.

FUNDING

Funding for open access charge: The European Union Seventh Programme under the FP7-PEOPLE-MCA-IRG funding scheme [256360 to E.Y.-L.].

Conflict of interest statement. None declared.

REFERENCES

- Schadt,E.E., Friend,S.H. and Shaywitz,D.A. (2009) A network view of disease and compound screening. *Nat. Rev. Drug Discov.*, **8**, 286–295.
- Vidal,M., Cusick,M.E. and Barabasi,A.L. (2011) Interactome networks and human disease. *Cell*, **144**, 986–998.
- Berggard,T., Linse,S. and James,P. (2007) Methods for the detection and analysis of protein-protein interactions. *Proteomics*, **7**, 2833–2842.
- Stark,C., Breitkreutz,B.J., Chatr-Aryamontri,A., Boucher,L., Oughtred,R., Livstone,M.S., Nixon,J., Van Auken,K., Wang,X., Shi,X. *et al.* (2011) The BioGRID Interaction Database: 2011 update. *Nucleic Acids Res.*, **39**, D698–D704.
- Salwinski,L., Miller,C.S., Smith,A.J., Pettit,F.K., Bowie,J.U. and Eisenberg,D. (2004) The database of interacting proteins: 2004 update. *Nucleic Acids Res.*, **32**, D449–D451.
- Aranda,B., Achuthan,P., Alam-Faruque,Y., Armean,I., Bridge,A., Derow,C., Feuermann,M., Ghanbarian,A.T., Kerrien,S., Khadake,J. *et al.* (2010) The IntAct molecular interaction database in 2010. *Nucleic Acids Res.*, **38**, D525–D531.
- Ceol,A., Chatr Aryamontri,A., Licata,L., Peluso,D., Briganti,L., Perfetto,L., Castagnoli,L. and Cesareni,G. (2010) MINT, the molecular interaction database: 2009 update. *Nucleic Acids Res.*, **38**, D532–D539.
- Su,A.I., Wiltshire,T., Batalov,S., Lapp,H., Ching,K.A., Block,D., Zhang,J., Soden,R., Hayakawa,M., Kreiman,G. *et al.* (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl Acad. Sci. USA*, **101**, 6062–6067.
- Ramskold,D., Wang,E.T., Burge,C.B. and Sandberg,R. (2009) An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. *PLoS Comput. Biol.*, **5**, e1000598.
- Bossi,A. and Lehner,B. (2009) Tissue specificity and the human protein interaction network. *Mol. Syst. Biol.*, **5**, 260.
- Emig,D., Kacprowski,T. and Albrecht,M. (2011) Measuring and analyzing tissue specificity of human genes and protein complexes. *EURASIP J. Bioinform. Syst. Biol.*, **2011**, 5.
- Safran,M., Dalah,I., Alexander,J., Rosen,N., Iny Stein,T., Shmoish,M., Nativ,N., Bahir,I., Doniger,T., Krug,H. *et al.* (2010) GeneCards Version 3: the human gene integrator. *Database (Oxford)*, **2010**, baq020.
- Lane,L., Argoud-Puy,G., Britan,A., Cusin,I., Duek,P.D., Evalet,O., Gateau,A., Gaudet,P., Gleizes,A., Masselot,A. *et al.* (2012) neXtProt: a knowledge platform for human proteins. *Nucleic Acids Res.*, **40**, D76–D83.
- Wu,C., Orozco,C., Boyer,J., Leglise,M., Goodale,J., Batalov,S., Hodge,C.L., Haase,J., Janes,J., Huss,J.W. III *et al.* (2009) BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol.*, **10**, R130.
- Yanai,I., Benjamin,H., Shmoish,M., Chalifa-Caspi,V., Shklar,M., Ophir,R., Bar-Even,A., Horn-Saban,S., Safran,M., Domany,E. *et al.* (2005) Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics*, **21**, 650–659.
- Berglund,L., Bjorling,E., Oksvold,P., Fagerberg,L., Asplund,A., Szigarty,C.A., Persson,A., Ottosson,J., Wernerus,H., Nilsson,P. *et al.* (2008) A gene-centric Human Protein Atlas for expression profiles based on antibodies. *Mol. Cell. Proteomics*, **7**, 2019–2027.
- Bradley,R.K., Merkin,J., Lambert,N.J. and Burge,C.B. (2012) Alternative splicing of RNA triplets is often regulated and accelerates proteome evolution. *PLoS Biol.*, **10**, e1001229.
- Guberman,J.M., Ai,J., Arnaiz,O., Baran,J., Blake,A., Baldock,R., Chelala,C., Croft,D., Cros,A., Cutts,R.J. *et al.* (2011) BioMart central portal: an open database network for the biological community. *Database (Oxford)*, **2011**, bar041.
- Lopes,C.T., Franz,M., Kazi,F., Donaldson,S.L., Morris,Q. and Bader,G.D. (2010) Cytoscape Web: an interactive web-based network browser. *Bioinformatics*, **26**, 2347–2348.
- Ashburner,M., Ball,C.A., Blake,J.A., Botstein,D., Butler,H., Cherry,J.M., Davis,A.P., Dolinski,K., Dwight,S.S., Eppig,J.T. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.
- Kuhn,M., Szklarczyk,D., Franceschini,A., von Mering,C., Jensen,L.J. and Bork,P. (2012) STITCH 3: zooming in on protein-chemical interactions. *Nucleic Acids Res.*, **40**, D876–D880.